Food and Chemical Toxicology 65 (2014) 269–279

Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

# Development of a consumer product ingredient database for chemical exposure screening and prioritization

M.-R. Goldsmith [a,*], C.M. Grulke [a], R.D. Brooks [b], T.R. Transue [c], Y.M. Tan [a], A. Frame [a,e], P.P. Egeghy [a], R. Edwards [d], D.T. Chang [a], R. Tornero-Velez [a], K. Isaacs [a], A. Wang [a,e], J. Johnson [a], K. Holm [a], M. Reich [f], J. Mitchell [g], D.A. Vallero [a], L. Phillips [a], M. Phillips [a], J.F. Wambaugh [a], R.S. Judson [a], T.J. Buckley [a], C.C. Dary [a]

[a] United States Environmental Protection Agency (U.S. EPA), Office of Research and Development, RTP, NC 27711, United States
[b] Student Services Contractor at U.S. EPA, RTP, NC, United States
[c] Lockheed-Martin Information Technology, RTP, NC 27711, United States
[d] North Carolina State University, 2200 Hillsborough St., Raleigh, NC 27695, United States
[e] Oak Ridge Institute for Science and Education Fellow, United States
[f] University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, United States
[g] Biosystems and Agricultural Engineering, Michigan State University, E. Lansing, MI 48824, United States

ARTICLE INFO

*Article history:*
Received 16 October 2013
Accepted 19 December 2013
Available online 27 December 2013

*Keywords:*
Chemical exposure
Consumer products
Ingredients
Product formulation
Near field exposure
Exposure prioritization

ABSTRACT

Consumer products are a primary source of chemical exposures, yet little structured information is available on the chemical ingredients of these products and the concentrations at which ingredients are present. To address this data gap, we created a database of chemicals in consumer products using product Material Safety Data Sheets (MSDSs) publicly provided by a large retailer. The resulting database represents 1797 unique chemicals mapped to 8921 consumer products and a hierarchy of 353 consumer product "use categories" within a total of 15 top-level categories. We examine the utility of this database and discuss ways in which it will support (i) exposure screening and prioritization, (ii) generic or framework formulations for several indoor/consumer product exposure modeling initiatives, (iii) candidate chemical selection for monitoring near field exposure from proximal sources, and (iv) as activity tracers or ubiquitous exposure sources using "chemical space" map analyses. Chemicals present at high concentrations and across multiple consumer products and use categories that hold high exposure potential are identified. Our database is publicly available to serve regulators, retailers, manufacturers, and the public for predictive screening of chemicals in new and existing consumer products on the basis of exposure and risk.

Published by Elsevier Ltd.

## 1. Introduction

The production and use of chemicals is a hallmark of a modern and consumer oriented society. Within this modern society, however, there is also a growing awareness that there are thousands of chemicals to which humans are unavoidably exposed every day (Glegg and Richards, 2007; Sanderson et al., 2013; Weschler, 2009; Anastas et al., 2010). Of particular interest are chemicals from anthropogenic sources that can be potentially controlled. However, tracking and controlling these chemicals is a daunting task since it has been estimated that there are some 80,000 different chemicals in commerce (NTP, 2002). To evaluate the potential risk to human health associated with the use of these chemicals, there are two primary considerations: the chemical hazard and exposure. Because of the large number of chemicals and limited information about hazard or exposure, there is a need for screening methods of evaluation that have minimal data requirements and

*Abbreviations:* ACToR, Aggregated Computational Toxicology Resource; ADME, Absorption, Distribution, Metabolism, and Elimination; CAS-RN, chemical abstracts service registry number; CBI, confidential business information; CGI, common gateway interface; CHAD, Consolidated Human Activity Database; CPCPdb, Consumer Product Chemical Profile Database; CPSC, Consumer Product Safety Commission; CTCP, Clinical Toxicology of Commercial Products; DSSTox, Distributed Structure-Searchable Toxicity Database; EFH, Exposure Factors Handbook; HPDB, Household Product Database; HTML, HyperText Markup Language; IUPAC, International Union of Pure and Applied Chemistry; MSDS, Material Safety Data Sheet; NACE, Nomenclature des Activités Économiques dans la Communauté Européenne; NAICS, North American Industry Classification System; NEISS, National Electronic Injury Surveillance System; NLM, National Library of Medicine; OCR, optical character recognition; OSHA, Occupational Safety and Health Administration; Perl, Practical Extraction and Reporting Language; PHP, P Hypertext Preprocessor; REACH, Registration, Evaluation, Authorisation and Restriction of Chemical substances; SHEDS, Stochastic Human Exposure and Dose Simulation; SPIN, Substances in Products in the Nordic Countries; SOP, standard operating procedure; TSCA, Toxic Substances Control Act; URL, Uniform Resource Locator.

* Corresponding author. Tel.: +1 919 541 0497.
*E-mail addresses:* goldsmith.rocky@epa.gov (M.-R. Goldsmith), tan.cecilia@epa.gov (Y.M. Tan), isaacs.kristin@epa.gov (K. Isaacs).

0278-6915/$ - see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.fct.2013.12.029

can accommodate hundreds or thousands of chemicals at time (i.e., high throughput). Under a multi-agency initiative (NRC, 2007), considerable progress has already been made with respect to the development and implementation of high-throughput hazard testing. Progress on toxicity testing has in part highlighted the need for complementary high throughput evaluation of exposure.

For the purpose of high throughput exposure evaluation, there is strong justification based on both measurement (Morgan et al., 2005; Pellizzari et al., 1999; Wallace, 1987) and modeling (Lai et al., 2000; Nazaroff et al., 2012) studies to consider first and foremost those chemicals that are found in consumer products. These chemicals are brought into our homes where circumstances exist for relatively high levels of exposure depending on a chemical's prevalence in products and the frequency of product use along with chemical's degradation or removal rates (Hertwich, 2005). Indeed, many chemical ingredients that are reactive outdoors become relatively persistent indoors in light of physically different substrate (e.g. carpet twill with numerous sorption sites) and decreased biodegradation and photodegradation rates. More open systems, such as the ambient atmosphere, enhance a chemical's decay rate compared to the more closed indoor systems (Vallero, 2010). The indoor residential environment is where people spend a large majority of time, is where susceptible individuals (very young and elderly and those who are sick or disabled) tend to spend even more of their time (Klepeis et al., 2001), and is where most consumer products tend to be stored and used. It is recognized that depending on their purpose, exposure to chemicals in consumer products can result through either direct or indirect routes of contact (Schettler, 2006; Rudel et al., 2003; Weschler and Nazaroff, 2008). Historical examples demonstrating both the exposure and health risk potential of chemicals in consumer products that have led to their subsequent mitigation include lead in paint and chlorpyrifos in household pesticides.

Because consumer product use is an important determinant for human exposure to a broad range of chemicals, it follows that information about the chemicals and their concentrations within those products is also an important consideration. There are a few potential sources for such information. One is Material Safety Data Sheets (MSDSs) which are required under the Occupational Safety and Health Administration (OSHA) Hazard Communication Standard to include product ingredients with known toxicity. They are intended to inform workers and emergency response personnel of hazards and their safe management. All hazardous components in excess of 1% (0.1% for carcinogens) are required to be disclosed through product labeling and the MSDS. OSHA defines a hazardous chemical very broadly as one that could possibly cause any physical or health effect under expected conditions of use or reasonably anticipated conditions of misuse. OSHA does not require MSDSs to be provided to consumers. However, with greater public interest for this kind of information, many manufacturers and retailers are providing MSDS as a public service. For example, Walmart has made the MSDS inventory of their products available online (http://msds.walmartstores.com) for more than 10 years. A second source of such data is the Consumer Product Safety Commission (CPSC). The CPSC has been collecting data of this nature for internal use and regulatory purposes for the last four decades (Bracken and Weiss, 1977; Byer et al., 1976). A third source is the publication Clinical Toxicology of Commercial Products (CTCP), which was one of the first to aggregate consumer product ingredient compendiums dating back three decades; it included 14,000 products and over 900 generic product formulations (Gosselin et al., 1984). More recent efforts by the National Library of Medicine (NLM) provide similar data known as the Household Product Database (HPDB) in a web-accessible format indexed by consumer product category. The licensed form of this database is the Consumer Product Information Database (CPID). It includes the MSDS listing and chemical percent composition by weight when available and is soft-linked to multiple NLM informatics resources (http://whatsinproducts.com/). This database is unique in that it provides quantitative composition information which is critical for evaluating the exposure potential (Jayjock et al., 2008). A fourth effort is underway by the Chemical Trade Associations that represent both manufacturers and formulators to develop a Consumer Product Ingredient Communication Initiative to provide consumers with information about ingredients in products (Egeghy et al., 2011).

Accordingly, the current research is motivated both by need and opportunity. The large and growing number of chemicals that are used in consumer products result in considerable exposure potential to complex mixtures; therefore, there is a need for high throughput evaluation of exposure as a fundamental component of risk screening. At the same time, public information is increasingly available on which chemicals comprise a product. What is lacking is the methodology for capturing the available data in a form that supports high throughput exposure screening for rapid risk analysis (Goldsmith et al., 2012). In this paper, we rely on available data sources for collecting a structured set of product compositions that can be used to inform aggregate exposure and to generate generic product formulations used in regulatory exposure assessment models (i.e. EFAST or RIVM's ConsExpo). This structured set of product composition we have collated is the Consumer Product Chemical Profiles database (CPCPdb). The development of this database is well aligned with the National Research Council's report on exposure science (NRC, 2012) espousing the efficient development of data and the application of computational methods to derive new models that better predict exposure. Not only is CPCPdb a valuable tool for quantifying exposures, to aid in risk assessments, it is a scientifically credible means for evaluating exposure potential for a large number of chemicals leading to more effective prioritization of hazard characterization resources. Rather than focusing first on the chemicals that are known to have toxic effects and asking if there is a likelihood of exposure (i.e. hazard-centric risk assessment), the CPCPdb can be used to identify the chemicals with which the population has greatest contact so *in vitro* and *in vivo* toxicity testing can be used more judiciously. Exposure-based prioritization can be used to focus hazard identification and assessment efforts to a subset of relevant chemicals, streamlining the risk assessment process by reducing the dimensionality of possible (i.e. all chemicals) chemical exposures to the probable set of chemicals requiring our immediate attention.

## 2. Methods

The methodology for building the CPCPdb can be broken down into three major steps:

1. Building and curating a database for consumer product ingredients and percent compositions using available MSDSs.
2. Identifying and annotating product use categories for all products in the database.
3. Evaluating data quality.

Descriptions of each of these efforts are contained below and captured in Fig. 1.

### *2.1. Building and curating a database for consumer product ingredient and percent compositions using available MSDSs*

#### *2.1.1. Identification, retrieval, extraction*

The first step necessary in the formation of the database was collecting consumer product MSDSs from a publicly available source. Walmart was chosen because (a) its consumer product portfolio is relatively diverse (e.g., personal care, automotive, arts & crafts, household pesticides, lawn/garden, cleaning, home maintenance, home improvement, office supplies, electronics); (b) Walmart's omnipresence in the consumer product marketplace is expected to provide a relatively accessible, high-market share product inventory with high consumer coverage; (c) its database of MSDS was available in Adobe PDF format; and (d) the documents were not only publically accessible, but also available without any visible